Chapter 5

Section 5.1

Check Your Understanding, page 292:

1. (a) If you asked a large sample of U.S. adults whether they usually eat breakfast, about 61% of them will answer yes.
(b) In a random sample of 100 adults, we would expect that around 61 of them will usually eat breakfast. However, the exact number will vary from sample to sample.
2. (a) This probability is 0. If an outcome can never occur, then it will occur in 0% of the trials.
(b) This probability is 1. If an outcome will occur on every trial, then it will occur in 100% of the trials.
(c) This probability is 0.01. An outcome that occurs in 1% of the trials is very unlikely, but will occur every once in a while.
(d) This probability is 0.6. An outcome that occurs in 60% of the trials will happen more than half of the time. Also, 0.6 is a better choice than 0.99 because the wording suggests that the event occurs often but not nearly every time.

Check Your Understanding, page 299:

1. Assign the members of the AP Statistics class the numbers 01-28 and the rest of the students numbers 29-95. Ignore the numbers 96-99 and 00. In Table D read off 4 two-digit numbers, making sure that the second number is different than the first and that the fourth number is different than the third. Record whether all four numbers are between 01 and 28 or not. Do this many times and compute the percent of times that all four students selected are in the AP Statistics class.
2. Assign the numbers 1-10 to Jeff Gordon, 11-40 to Dale Earnhardt, Jr., 41-60 to Tony Stewart, 61-85 to Danica Patrick, and 86-100 to Jimmie Johnson. Do RandInt(1,100) until you have at least one number (box) for each of the 5 drivers. Count how many numbers (boxes) you had to select in order to get at least one of each driver. Do this many times and compute the percent of times that it takes 23 or more boxes to get the full set of drivers.

Exercises, page 300:

5.1 (a) If we use a polygraph machine on many, many people who are all telling the truth, about 8% of the time, the machine will say that the people are lying.
(b) Answers will vary. A false positive would mean that a person telling the truth would be found to be lying. A false negative would mean that a person lying would be found to be telling the truth. The U.S. judicial system is set up to think that a false positive would be worse – that is, saying that someone is guilty (lying) who is innocent is worse than finding someone to be truthful when they are guilty (lying).

5.2 (a) If we test many, many women with breast cancer, about 10% of the time the test will indicate that the woman does not have breast cancer.
(b) A false negative is a more serious error because a woman with breast cancer will not get potentially life-saving treatment. A false positive would result in temporary stress until a more thorough examination is performed.

5.3 (a) If we look at many families where the husband and wife both carry this gene, approximately 25% of the families will have a first-born child that develops cystic fibrosis.

(b) No. Although we expect about 1 in every 4 children in families like this to develop cystic fibrosis in the long run, it is possible that the number of children with cystic fibrosis will be smaller or larger by random chance in a family with four children.

5.4 (a) If we look at many, many hands of poker in which you hold a pair, you will make four of a kind about 8.8% of the time.
(b) No. Although we expect about 8.8% of hands with a pair to get four of a kind in the long run, it is possible that the number of hands that get four of a kind will be smaller or larger by random chance in a sample of 1000 such hands.

5.5 (a) Answers will vary. For example, on one set of 25 spins, we obtained 16 tails and 9 heads. Based on this set of 25 spins the probability of heads is approximately $\frac{9}{25} = 0.36$.
(b) You could get an even better estimate by spinning the coin many more times.

5.6 (a) Answers will vary. For example, on one set of 25 observations, we obtained 11 tails and 14 heads. Based on this set of 25 observations the probability of heads is approximately $\frac{14}{25} = 0.56$.
(b) You could get an even better estimate by knocking down the coin many more times.

5.7 In the short run, there was quite a bit of variability in the percentage of made free throws. However, the percentage of made free throws became less variable and approached 0.30 as the number of shots increased.

5.8 In the short run, there was lots of variability in the proportion of heads. In the long run, this proportion became less variable and settled down around 0.50 for both sets of 5000 tosses.

5.9 No, the TV commentator is incorrectly applying the law of large numbers to a small number of at-bats for the player.

5.10 No, the weather in one year might be related to the weather in previous years. Plus the TV weather man is incorrectly applying the law of large numbers to a small number of years.

5.11 (a) There are 10,000 four-digit numbers (0000, 0001, 0002,…, 9999), and each is equally likely to be chosen. One way to see this is to consider writing each of the numbers on a slip of paper and putting all of the numbers in a big box. Then you randomly select any slip of paper. Each of the 10,000 slips in the box, including the ones with 2873 and 9999 on them, is equally likely to be the one you select.
(b) Most people would say that 2873 is more likely than 9999 to be randomly chosen. To many it somehow "looks" more random – we don't "expect" to get the same number four times in a row. It would be best to choose a number that others would avoid so you don't have to split the pot with as many other people if you win.

5.12 (a) The wheel is not affected by its past outcomes—it has no memory. So on any one spin, black and red remain equally likely.
(b) The gambler is wrong again. Removing a card changes the composition of the remaining deck. If you hold 5 red cards, the deck now contains 5 fewer red cards, so the probability of being dealt another red card decreases.

5.13 (a) Let diamonds, spades, and clubs represent making a free throw and hearts represent missing. Deal one card from the deck.
(b) Let the two-digit numbers 00-74 represent making the free throw and 75-99 represent missing. Read a two-digit number from Table D.
(c) Let 1-3 represent the player making the free throw and 4 represent a miss. Generate a random integer from 1-4.

5.14 (a) Let 1 and 2 represent a green light and 3-6 represent a red light. Roll the die once.
(b) Let the numbers 1, 2 and 3 represent a green light and 4-9 represent a red light. Ignore 0. Look up one number in the table.
(c) Let 1 represent a green light and 2 and 3 represent a red light. Generate a random integer from 1-3.

5.15 (a) There are 19 numbers between 00 and 18, 19 numbers between 19 and 37, and 3 numbers between 38 and 40. This assignment of numbers changes the probabilities of the three different outcomes.
(b) There is no reason to skip numbers that have already been encountered in the table. For example, if the first number selected was 08, and you decide to skip repeated numbers, then the probability of selecting a left-hander on the next selection would be 9% instead of 10%.

5.16 (a) There is no reason to skip numbers that have already been encountered in the table. For example, if the first number selected was 18, and you decide to skip repeated numbers, then the probability of selecting an obese adult on the next selection would be 35/99 = 35.35% instead of 36%.
(b) This will give the numbers 0 through 9 an equal chance of occurring. However, if boys and girls are equally likely, there are more ways to have 4, 5, or 6 boys than 0 or 9 boys.

5.17 (a) This is a legitimate simulation. The chance of rolling a 1, 2 or 3 is 75% on a 4-sided die and the 100 rolls represent the 100 randomly selected U.S. adults.
(b) This is not a valid design because the probability of heads is 50% (assuming the coin is fair) rather than 60%. This method will underestimate the number of times she hits the center of the target.

5.18 (a) This is not a valid design because you are not putting the card back in the deck after dealing. For example, if the first card was red, and you do not replace it in the deck before drawing the next card, then there is only a 25/51 (49%) probability that the next person selected will be addicted to texting.
(b) This is an appropriate design because there is a 95% probability of getting a number between 00 and 94 and the five pairs of digits represent the five serve attempts.

5.19 (a) What is the probability that, in a random selection of 10 passengers, none from first class are chosen?
(b) Number the first class passengers 01-12 and the other passengers 13-76. Ignore all other numbers. Look up two-digit numbers in Table D until you have 10 unique numbers (no repetitions because you do not want to select the same person twice). Count the number of two-digit numbers between 01 and 12.
(c) The numbers read in pairs are: **71 48 70** 99 84 **29** ***07*** 71 48 **63 61 68 34** 70 **52**. The bold numbers indicate people who have been selected. The other numbers are either too large (over 76) or have already been selected. There is one person among the 10 selected who is in first class in this sample (in italics).

(d) Because no first class passenger was chosen in 15% of the samples, it seems plausible that the actual selection was random.

5.20 (a) What is the probability that, in selecting 7 tiles from 100, all 7 are vowels?
(b) Let the numbers 01-42 represent the vowels, 43-98 represent the consonants, and 99 and 00 represent the blank tiles. Look up two-digit numbers in Table D until you have 7 unique numbers (no repetitions since once you pull one tile from the bag you cannot pull it again). Record whether all 7 numbers are between 01 and 42 or not.
(c) The numbers read in pairs are: 00 69 **40** 59 77 **19** 66. The two numbers in bold are vowels, so the whole sample is not just vowels.
(d) Because getting a sample with all vowels only occurred in 2 trials out of 1000, it is very unlikely to get a sample like this by chance.

5.21 (a) Use a random integer generator to select 30 numbers from 1 to 365. Record whether or not there were any repeats in the sample.
(b) Answers will vary. One possible answer: We used RandInt(1,365,30) to perform 5 repetitions. There were repeats in all 5 samples.
(c) Answers will vary. One possible answer: After the simulation we would not be surprised that the probability is 0.71 since we found repeats in 100% of our samples. Before the simulation this may seem surprising since there are 365 different days that people could be born on and we only selected 30 people.

5.22 (a) Answers will vary. One possible answer: In our sample we won 11 out of 25 times when we stayed and we won 19 out of 25 times when we switched.
(b) Answers will vary. One possible answer: We agree with Marilyn—we won much more often when we switched doors.

5.23 (a) In the simulation, 43 of the 200 samples yielded a sample percentage of at least 55%. Obtaining a sample percentage of 55% or higher is not particularly unusual when 50% of all students recycle.
(b) Only 1 of the 200 samples yielded a sample percentage of at least 63%. This means that if 50% of all students recycle, we would see a sample percentage of at least 63% in only about 0.5% of samples. Because getting a sample percentage of at least 63% is very unlikely, we have convincing evidence that the percentage of all students who recycle is larger than 50%.

5.24 (a) If 27 out of 60 say they leave the water running, this means that 45% of the sample leave the water running. In the simulation, 44 of the 200 samples showed 45% or fewer leaving the water running. Obtaining a sample percentage of 45% or lower is not particularly unusual when 50% of all people brush with the water off.
(b) However, if 18 out of 60 say they leave the water running, this means that 30% of the sample leave the water running. In the simulation, no samples had a percentage this low. Because obtaining a sample percentage of at most 30% is very unlikely, we have strong evidence that fewer than 50% of the school's students brush their teeth with the water off.

5.25 *State*: What is the probability that, in a sample of 4 randomly selected U.S. adult males, at least one of them is red-green colorblind? *Plan*: We'll use Table D to simulate selecting samples of 4 men. Because 7% of men are red-green colorblind, let 00-06 denote a colorblind man and 07-99 denote a non-colorblind man. Read 4 two-digit numbers from the table for each sample and record whether or not the sample had at least one red-green colorblind man in it. *Do*: We did 50 repetitions. We started on row 109 from Table D. The first 5 samples are listed here: Sample 1: 36 00 91 93 (1 colorblind); Sample 2: 65 15 41 23 (no colorblind); Sample 3: 96 38 85 45 (no

colorblind); Sample 4: 34 68 16 83 (no colorblind); Sample 5: 48 54 19 79 (no colorblind). In our 50 samples, 15 had at least one colorblind man in them. *Conclude*: Based on our 50 simulated samples, the probability that a sample of 4 men would have at least one colorblind man is approximately 15/50 = 0.30.

5.26 *State*: What is the probability that none of the winning numbers will appear on a Lotto ticket with six numbers randomly selected? *Plan*: Pick the 6 numbers from 1-49 to be the "winning" numbers. We'll pick 1-6 but it could be any 6 numbers. Then, using a random integer generator, pick 6 numbers from 1-49 (ignoring repeats). Record how many numbers are 1-6. *Do*: We did 50 repetitions. The first five tickets were: ticket 1: 32 21 49 37 18 44 (0 matches); ticket 2: 23 19 10 33 44 31 (0 matches); ticket 3: 36 41 48 40 19 **3** (1 match); ticket 4: 43 25 33 **6** 40 8 (1 match); ticket 5: 25 11 15 42 27 **6** (1 match). In our 50 samples, 27 had no matches. *Conclude*: Based on our 50 simulated samples, the probability of getting no matches on a lottery ticket is approximately 27/50 = 0.54. The lottery player should not be surprised as this probability is fairly large.

5.27 *State*: What is the probability that it takes 20 or more selections in order to find one man who is red-green colorblind? *Plan*: We'll label the numbers 0–6 as colorblind men and 7–99 as non-colorblind men. Use technology to pick integers from 0 to 99 until we get a number between 0 and 6. Count how many numbers there are in the sample. *Do*: We did 50 repetitions. The first repetition is given here: 17 33 49 41 **2**. For this repetition we chose 5 men in order to get the one colorblind man (in bold). In 16 of our 50 samples, it took 20 or more selections to get one colorblind man. *Conclude*: Based on our 50 simulated samples, the probability of needing 20 or more selections to get one colorblind man is approximately 16/50 = 0.32. We should not be surprised as this probability is fairly large.

5.28 *State*: What is the probability that it takes 30 or more games to have a player bingo on the first turn? *Plan*: We'll label the numbers 0–2 as a bingo and 3–99 as not a bingo. Use technology to pick integers from 0 to 99 until we get a number between 0 and 2. Count how many numbers there are in the sample. *Do*: We did 50 repetitions. The first repetition is given here: 14 33 12 10 5 42 95 81 72 21 31 55 74 79 70 23 63 17 65 62 81 11 92 97 7 38 8 33 29 5 6 85 28 63 44 46 99 18 47 52 48 6 36 88 83 **1**. For this repetition it took 46 games for someone to bingo on the first turn. In 21 of our 50 repetitions, it took 30 or more games to get a bingo on the first turn. *Conclude*: Based on our 50 simulated games, the probability of needing 30 or more games to get a bingo on the first turn is approximately 21/50 = 0.42. We should not be surprised as this probability is fairly large.

5.29 *State*: What is the probability that a random assignment of 8 men and 12 women to two groups of 10 each will result in at least 6 men in the same group? *Plan*: Number the men 1–8 and the women 9–20. Use technology to pick 10 unique integers between 1 and 20 for one group, the remaining 10 numbers are in the other group. Record whether at least 6 of the numbers between 1 and 8 are in one group. *Do*: We did 50 repetitions. The first repetition resulted in Group 1: 17 **05 04** 14 11 10 **03** 15 **08 01**; Group 2: **02** 19 12 **06** 18 09 16 13 **07** 20. The men are shown in bold. There were 5 men in the first group and three men in the second group, so neither group had 6 or more men. In our 50 repetitions, 9 had one group with 6 or more men in it. *Conclude*: Based on our 50 simulated random assignments, the probability of getting 6 or more men in one group is approximately 9/50 = 0.18. They should not be surprised as this probability is fairly large.

5.30 *State*: What is the probability that the train will be late on at least 2 out of 6 randomly selected days if the train arrives on time on 90% of days? *Plan*: Let the digits 1-9 represent days

that the train arrives on time and the digit 0 represent days when the train does not arrive on time. Use Table D to pick 6 digits and record whether there were at least two 0's among the 6 digits. *Do*: We did 50 repetitions starting on line 122. The first 5 repetitions are as follows: 138738 (no); 159895 (no); **0**529**0**9 (yes); **0**87359 (no); 275186 (no). In 5 of our 50 repetitions, the train arrived late on 2 or more of the 6 days. *Conclude*: Based on our 50 repetitions, the probability of having the train arrive late twice or more in 6 days is approximately 5/50 = 0.10. Because this is not a very small probability, we do not have convincing evidence that the company's claim isn't true.

5.31 c

5.32 a

5.33 b

5.34 c

5.35 c

5.36 b

5.37 (a) The population is adult U.S. residents with telephones and the sample is the 353,564 adults who were interviewed.
(b) Because the interviews were conducted using the telephone, those people who do not have a telephone were excluded. In general, people who do not have phones tend to be poorer and therefore may experience more stress in their lives than the population as a whole. This would lead to an underestimate of the proportion in the population who experienced stress a lot of the day yesterday.

5.38 (a) Both distributions are skewed to the right with the drivers generally taking longer to leave when someone is waiting for the space. The median time to leave was about 40 seconds for the drivers with no one waiting and about 47 seconds for the drivers with someone waiting. There is more variability for the drivers with someone waiting as well – they had an *IQR* of 23.03 seconds as opposed to an *IQR* of 12.87 seconds for those with no one waiting. Finally there were no outliers for those with someone waiting, but there were two high outliers for those with no one waiting. (b) Not necessarily. The researchers merely observed what was happening, they did not randomly assign the treatments of either having a person waiting or not to the drivers of the cars leaving the lot.